Claims

1.   A method for using machine learning to predict an outcome comprising the steps of:

applying a test subset of a data set to a machine learning system, wherein the step of applying a test subset of data includes applying a first subset of data corresponding to a first outcome, applying a second subset of data corresponding to a second outcome and consisting of a set of nearest neighbors to the first outcome, and applying a third subset of data corresponding to the second outcome; and

using a plurality of machine learning methods to develop one or more sets of rules to predict results from the subset.

2.   The method of claim 1, wherein the step of applying a third subset of data involves applying a third subset of data that does not consist of nearest neighbors to the first outcome.

3.   The method of claim 2, wherein the step of applying a third subset of data includes randomly selecting a subset of the test subset of data corresponding to the second outcome and not consisting of nearest neighbors to the first outcome as the third subset.

4.   The method of claim 1, wherein the step of applying a test subset of data includes applying a test subset of data wherein the test subset of data includes records having an outcome variable and a plurality of feature variables.

5.   The method of claim 4, further comprising the step of identifying a set of nearest neighbors using values of the outcome variable for the test subset of data.

6.   The method of claim 5, wherein the step of applying a second subset of data includes randomly selecting a subset of the identified set of nearest neighbors as the second subset.

7.   The method of claim 5, wherein the step of applying a second subset of data includes selecting as the second subset all of the identified set of nearest neighbors.

8.   The method of claim 4, further comprising the step of identifying a set of nearest neighbors using values of the plurality of feature variables for the test subset of data.

9.   The method of claim 1, further comprising the step of validating the one or more sets of rules using the data set.

10.   The method of claim 9, wherein the step of validating the one or more sets of rules includes obtaining one or more accuracy measures for the rules using the data set.

11.   The method of claim 9, wherein the step of validating the one or more sets of rules includes obtaining one or more accuracy measures for the rules using the entire data set.

12.   The method of claim 11, wherein the step of validating the one or more sets of rules further includes obtaining the one or more accuracy measures for the test subset of the data set.

13.   The method of claim 11, wherein the step of obtaining one or more accuracy measures includes obtaining measures of the positive predictive value, the negative predictive value, the sensitivity, and the selectivity of the rules.

14.   The method of claim 1, wherein the step of using a plurality of machine learning methods includes developing a set of interim rules using the plurality of machine learning methods, evaluating the set of interim rules, and developing a revised set of interim rules using the results of the evaluating step.

15.   The method of claim 14, wherein the step of evaluating the set of interim rules includes applying a user-selectable fitness function.

16.   The method of claim 14, wherein the step of evaluating the set of interim rules includes applying a fitness function based on one or more of the sensitivity, the positive predictive value, and the correlation coefficient of the interim rules.

17.   A method for using machine learning to predict results comprising the steps of:

applying a representation of a subset of a data set to a machine learning system;

repeating for a plurality of cycles:

using a plurality of machine learning methods to develop a set of rules from the applied representation of the data;

evaluating the set of rules using a user-selectable fitness function; and

modifying the machine learning methods using the results of the evaluating step; and

presenting a final set of rules.

18.   The method of claim 17, wherein the step of evaluating a set of rules includes using a user-selectable fitness function based on one or more of: the number of true positives, the

number of true negatives, the number of false positives, and the number of false negatives that the set of rules obtains from the subset of the data set.

19.   The method of claim 17, wherein the step of evaluating a set of rules includes using a user-selectable fitness function based on the sensitivity and the positive predictive value of the rules.

20.   The method of claim 17, wherein the step of evaluating a set of rules includes using a user-selectable fitness function based on the sensitivity, the positive predictive value, and the correlation coefficient of the rules.

21.   The method of claim 17, further comprising, in at least one of the plurality of cycles, developing one or more new representations of the data for use by the plurality of machine learning methods in a subsequent cycle.

22.   A method for using machine learning to predict a positive or a negative outcome, where the positive outcome is less likely than the negative outcome, the method comprising the steps of:

applying a test subset of a data set to a machine learning system, wherein the step of applying a test subset of data includes applying a first subset of data corresponding to a first outcome, applying a second subset of data corresponding to a second outcome and consisting of a set of nearest neighbors to the first outcome, and applying a third subset of data corresponding to the second outcome; and

using a plurality of machine learning methods to develop one or more sets of rules to predict an outcome from the subset.

23.   The method of claim 22, wherein the step of using a plurality of machine learning methods to develop one or more sets of rules includes applying a user-selectable fitness function to develop the one or more sets of rules.